数理解析研究所講究録
1043 巻 1998 年 169-177

Controlled Markov Set-Chains Under Average Criteria

Jinjie Song*
Masanori Hosaka* and Masami Kurano**
* Graduate School of Science and Technology,
** Faculty of Education,
Chiba University
Yayoi-cho, Inage-ku, Chiba, Japan, 263

Abstract

In a controlled Markov set-chain with finite state and action spaces, we find a policy, called average-optimal, which maximizes Cesaro sums of each time's reward over all stationaly policies under some partial order.

Analysing the behavior of expected total rewards over the T-horizon as T approaches ∞ under irreducibility condition, the average rewards from any stationary policy are characterized.

Also, we investigate the left and right side optimality equations, by which the existence of an average-optimal policy is shown. A numerical example is given.

Keywords: Controlled Markov set-chains, average reward criterion, a interval arithemetic.

AMS 1991 subject classification. Primary: 90c40; Secondary: 90c39.

1 Introduction and notations

In standard Markov decision processes [cf.2,7,8,12], we treat the case that the transition probability of the state varies in some given domain at each time and its variation is unknown or unobservable.

For the sake of analysing such a case, Kurano et al [10] has introduced a new decision model, called a controlled Markov set-chain, based on Markov set-chains [3,4,5,6], and discussed the optimization of the discounted expected rewards under some partial order.

In our previous paper [14], we have tried to find a policy, called average-optimal, which maximizes Cesaro sums of each time's reward under some partial order.

The main condition the authors imposed on was of uniformly scrambling type, under which the dynamic programming operator for our model became a contraction in a span seminorm.

The objective of this paper is to prove the same redults as [14] without uniformly scrambling condition Analysing the behavior of the expected total rewards over the T-horizon as T approches ∞ under irreducibility condition, the average rewards from any stationary policy are characterized.

Also, we investigate the left and right side optimality equations, by which the existence of an average-optimal policy is shown. Our proof is done by applying the results of Bather [1] and the idea of policy improvement (cf.[8]). The proofs are omitted for the sake of shrinking the page.

We adopt the notations in [6,10,11].

Let R, R^n and $R^{n\times m}$ be the sets of real numbers, real n- dimensional column vectors and real $n \times m$ matrices, respectively. We shall identify $n \times 1$ matrices with vectors and 1×1 matrices with real numbers, so that $R = R^{1\times 1}$ and $R^n = R^{n\times 1}$. Also, we denote by R_+, R^n_+ and $R^{n\times m}_+$ the subsets of entrywise non-negative elements in R, R^n and $R^{n\times m}$, respectively.

We equip $R^{n\times m}$ with the componentwise relations $\leq, <, \geq, >$. For any $\underline{A} = (\underline{a}_{ij})$, $\overline{A} = (\overline{a}_{ij})$ in $R^{n\times m}_+$ with $\underline{A} \leq \overline{A}$, we define the set of stochastic matrices, $\langle \underline{A}, \overline{A}\rangle$, by

$$\langle \underline{A}, \overline{A}\rangle := \{A \mid A = (a_{ij}) \text{ is an } n \times m \text{ stochastic matrix with } \underline{A} \leq A \leq \overline{A}\},$$

Let

$$\mathcal{M}_n := \{\mathcal{A} = \langle \underline{A}, \overline{A}\rangle \mid \langle \underline{A}, \overline{A}\rangle \neq \emptyset, \underline{A} \leq \overline{A} \text{ and } \underline{A}, \overline{A} \in R^{n\times n}_+\}.$$

The product of $\mathcal{A}$ and $\mathcal{B} \in \mathcal{M}_n$ is defined by

$$\mathcal{A}\mathcal{B} := \{AB \mid A \in \mathcal{A}, B \in \mathcal{B}\}.$$

For any sequence $\{\mathcal{A}_i\}_{i=1}^{\infty}$ with $\mathcal{A}_i \in \mathcal{M}_n$ $(i \geq 1)$, we define the multiproduct inductively by

$$\mathcal{A}_1\mathcal{A}_2\cdots\mathcal{A}_k := (\mathcal{A}_1\cdots\mathcal{A}_{k-1})\mathcal{A}_k \ (k \geq 2).$$

Denote by $C(R_+)$ the set of all bounded and closed intervals in R_+.

Let $C(R_+)^n$ be the set of all n-dimemsional column vectors whose elements are in $C(R_+)$, i.e.,

$$C(R_+)^n := \{D = (D_1, D_2, \cdots, D_n)' \mid D_i \in C(R_+) \ (1 \leq i \leq n)\}.$$

where d' denotes the transpose of a vector d.

The following arithemetics are used in Section 2.

For $D = (D_1, D_2, \cdots, D_n)'$, $E = (E_1, E_2, \cdots, E_n)' \in C(R_+)^n$, $h \in R^n_+$ and $\lambda \in R_+$, $D + E = \{d + e \mid d \in D, e \in E\}$, $\lambda D = \{\lambda d \mid d \in D\}$ and $h + D = \{h + d \mid d \in D\}$.

If $D = ([\underline{d}_1, \overline{d}_1], \cdots, [\underline{d}_n, \overline{d}_n])'$, D will be denoted by $D = [\underline{d}, \overline{d}]$, where $\underline{d} = (\underline{d}_1, \cdots, \underline{d}_n)'$, $\overline{d} = (\overline{d}_1, \cdots, \overline{d}_n)'$ and $[\underline{d}, \overline{d}] = \{d \mid d \in R^n_+, \underline{d} \leq d \leq \overline{d}\}$.

For any $D = (D_1, D_2, \cdots, D_n)' \in C(R_+)^n$ and subset G of $R^{1\times n}_+$ the product of G and D is defined as

$$GD = \{gd \mid g = (g_1, \cdots, g_n) \in G, \ d = (d_1, \cdots, d_n)' \in D, \ d_i \in D_i \ (1 \leq i \leq n)\}.$$

The following results are used in the sequel.

Lemma 1.1 ([3,10])

(i) Any $\mathcal{A} \in \mathcal{M}_n$ is a convex polytope in the vector space $R^{n\times n}$.

(ii) For any compact convex subset $G \subset R_+^{1\times n}$ and $D = (D_1, D_2, \cdots, D_n)' \in C(R_+)^n$, it holds $GD \in C(R_+)$.

We will give a partial order $\geq, >$ on $C(R_+)$ by the definition :
For $[c_1, c_2]$, $[d_1, d_2] \in C(R_+)$,

$$[c_1, c_2] \geq [d_1, d_2] \text{ if } c_1 \geq d_1, \; c_2 \geq d_2,$$

and

$$[c_1, c_2] > [d_1, d_2] \text{ if } [c_1, c_2] \geq [d_1, d_2] \text{ and } [c_1, c_2] \neq [d_1, d_2].$$

For $v = (v_1, v_2, \cdots, v_n)'$ and $w = (w_1, w_2, \cdots, w_n)' \in C(R_+)^n$, we write $v \geq w$ if $v_i \geq w_i$, $1 \leq i \leq n$ and $v > w$ if $v \geq w$ and $v \neq w$.
Define a metric Δ on $C(R_+)^n$ by

$$\Delta(v, w) := \max_{i \in S} \delta(v_i, w_i)$$

for $v = (v_1, v_2, \cdots, v_n)'$, $w = (w_1, w_2, \cdots, w_n)' \in C(R_+)^n$, where δ is the Hausdorff metric on $C(R_+)$ and given by

$$\delta([a, b], [c, d]) := |a - c| \vee |b - d| \text{ for } [a, b], [c, d] \in C(R_+),$$

where $x \vee y = \max\{x, y\}$.

Obviously, $(C(R_+)^n, \Delta)$ is a complete metric space.

A controlled Markov set-chain consists of four object; S, A, $\underline{q}$, $\overline{q}$, r, where $S = \{1, 2, \cdots, n\}$ and $A = \{1, 2, \cdots, k\}$ are finite sets and for each $(i, a) \in S \times A$, $\underline{q} = \underline{q}(\cdot|i, a) \in R_+^{1\times n}$ $\overline{q} = \overline{q}(\cdot|i, a) \in R_+^{1\times n}$ with $\underline{q} \leq \overline{q}$ and $\langle \underline{q}, \overline{q} \rangle \neq \emptyset$ and $r = r(i, a)$ a function on $S \times A$ with $r \geq 0$. Note that A is used as the set in this section, different from that in the previous section. We interpret S as the set of states of some system, and A as the set of actions available at each state.

When the system is in state $i \in S$ and we take action $a \in A$, we move to a new state $j \in S$ selected according to the probability distribution on S, $q(\cdot|i, a)$, and we recieve a return $r(i, a)$, where we know only that $q(\cdot|i, a)$ is arbitrarily chosen from $\langle \underline{q}(\cdot|i, a), \overline{q}(\cdot|i, a) \rangle$. This process is then repeated from the new state j.

Denote by F the set of functions from S to A.

A policy π is a sequence $(f_1, f_2, \cdots)$ of functions with $f_t \in F$, $(t \geq 1)$. Let Π denote the class of policies.

We denote by f^∞ the policy $(h_1, h_2, \cdots)$ with $h_t = f$ for all $t \geq 1$ and some $f \in F$. Such a policy is called stationary, denoted simply by f, and the set of stationary policies is denoted by Π_F.

We associate with each $f \in F$ the n-dimensional column vector $r(f) \in R_+^n$ whose ith element is $r(i, f(i))$ and the set of stochastic matrices $\mathcal{Q}(f) := \langle \underline{Q}(f), \overline{Q}(f) \rangle \in \mathcal{M}_n$, where the (i, j) elements of $\underline{Q}(f)$ and $\overline{Q}(f)$ are $\underline{q}(j|i, f(i))$ and $\overline{q}(j|i, f(i))$, respectively, and $\langle \underline{Q}(f), \overline{Q}(f) \rangle$ is already defined.

For any $\pi = (f_1, f_2, \cdots) \in \Pi$, let $v_1(\pi) = r(f_1)$ and

$$v_T(\pi) = \{r(f_1) + Q_1 r(f_2) + \cdots + Q_1 Q_2 \cdots Q_{T-1} r(f_T)$$

$$| Q_i \in \mathcal{Q}(f_i), i = 1, 2, \cdots, T-1\} \quad (T \geq 2). \tag{1.1}$$

We observe, for example that

$$v_3(\pi) = r(f_1) + \mathcal{Q}(f_1)\left(r(f_2) + \mathcal{Q}(f_2) r(f_3)\right),$$

so that by Lemma 1.1 (ii) $v_T(\pi) \in C(R_+)^n$ for all $T \geq 1$.

For any $\pi \in \Pi$, let

$$v(\pi) := \liminf_{T\to\infty} \frac{1}{T} v_T(\pi), \tag{1.2}$$

where, for a sequence $\{D_k\} \subset C(R_+)^n$,

$$\liminf_{k\to\infty} D_k := \left\{ x \in R^n \,\middle|\, \limsup_{k\to\infty} \delta_1(x, D_k) = 0 \right\},$$

and $\delta_1(x, D) = \inf_{y\in D} \delta_2(x, y)$, δ_2 is a metric in R^n. Since $v(\pi) \in C(R_+)^n, v(\pi)$ is written as $v(\pi) = [\underline{v}(\pi), \overline{v}(\pi)]$.

Definition A policy $f^* \in \Pi_F$ is called *average-optimal* if there does not exist $f \in \Pi_F$ such that $v(f^*) < v(f)$.

In the above definition, we confine ourselves to the stationary policies, which simplifies our discussion in the sequel.

In Section 2, irreducibility condition for the class of transition matrices is introduced, under which the interval equations concerning the average rewards are investigated.

In Section 3, the asymptotic behavior of $v_T(f)$ as T approaches ∞ is obtained. And in Section 4, the left and right side optimality equations are given and the existence of an average-optimal policy is proved.

2 Assumption and preliminary lemmas

Heceforth, the following assumption will remain operative.

Assumption A (irreducibility) For any $f \in F$, each $Q \in \mathcal{Q}(f)$ is irreducible, i.e., $Q^t > 0$ for some $t \geq 1$.

Obviously, if $\underline{Q}(f)$ is irreducible in the sence of non-negative matrix (cf.[13]), Assumption A holds.

The following facts about Markov matrices are well-known (cf.[2,9]).

Lemma 2.1 For any $f \in F$, let Q be any matrix in $\mathcal{Q}(f)$.

(i) The sequence $(I + Q + \cdots + Q^t)/(t+1)$ converges as $t \to \infty$ to a stochastic matrix Q^* with $Q^*Q = Q^*$, $Q^* > 0$ and rank$(Q^*) = 1$.

(ii) The matrix Q^* in (i) is uniquely determind by $Q^*Q = Q$ and rank$(Q^*) = 1$.

Associated with each $f \in F$ is a corresponding operator $L(f)$, mapping $C(R_+)^n$ into $C(R_+)^n$, defined as follows.

For $v \in C(R_+)^n$,

$$L(f)v := r(f) + \mathcal{Q}(f)v. \tag{2.1}$$

Note that from Lemma 1.1, $L(f)v \in C(R_+)^n$ for each $v \in C(R_+)^n$.

Putting $v = [\underline{v}, \overline{v}]$ with $\underline{v} \leq \overline{v}$, $\underline{v}, \overline{v} \in R_+^n$, (2.1) can be written as

$$L(f)v = [\underline{L}(f)\underline{v}, \overline{L}(f)\overline{v}], \tag{2.2}$$

where $\underline{L}$ and $\overline{L}$ are operators, mapping R^n into R^n, defined by :

$$\underline{L}(f)v = r(f) + \min_{Q \in \mathcal{Q}(f)} Qv, \tag{2.3}$$

$$\overline{L}(f)v = r(f) + \max_{Q \in \mathcal{Q}(f)} Qv. \tag{2.4}$$

and min (max) represents componentwise minimization (maximization).

Let $\mathrm{e} = (1, 1, \cdots, 1)'$.

Here, for any $f \in F$, we consider the interval equation

$$r(f) + \mathcal{Q}(f)\boldsymbol{h} = \boldsymbol{v} + \boldsymbol{h}, \tag{2.5}$$

where $\boldsymbol{v} = [\underline{v}\mathrm{e}, \overline{v}\mathrm{e}]$, $\boldsymbol{h} = [\underline{h}, \overline{h}] \in C(R)^n$, $\underline{v}$, $\overline{v} \in R$, $\underline{h}$, $\overline{h} \in R^n$ with $\underline{v} \leq \overline{v}$, $\underline{h} \leq \overline{h}$.

Obviously, (2.5) can be rewritten by

$$r(f) + \min_{Q \in \mathcal{Q}(f)} Q\underline{h} = \underline{v}\mathrm{e} + \underline{h} \tag{2.6}$$

$$r(f) + \max_{Q \in \mathcal{Q}(f)} Q\overline{h} = \overline{v}\mathrm{e} + \overline{h} \tag{2.7}$$

$$\text{where } \underline{v} \text{ , } \overline{v} \in R \text{ , } \underline{h} \text{ , } \overline{h} \in R^n \text{ with } \underline{v} \leq \overline{v} \text{ , } \underline{h} \leq \overline{h}. \tag{2.8}$$

Then, by a slight modification of the proof of Theorem 2.4 in Bather [1], we have the following lemma.

Lemma 2.2 ([1]) The interval equation (2.5) has a solution.

For simplicity of the notation, let, for any $d \in R^n$ and $f \in F$,

$$\underline{\mathcal{Q}}(f, d) = \left\{ Q \in \mathcal{Q}(f) \,\middle|\, Qd = \min_{Q \in \mathcal{Q}(f)} Qd \right\},$$

and

$$\overline{\mathcal{Q}}(f, d) = \left\{ Q \in \mathcal{Q}(f) \,\middle|\, Qd = \max_{Q \in \mathcal{Q}(f)} Qd \right\}.$$

Lemma 2.3 For any $f \in F$, the interval equation (2.5) determines $\boldsymbol{v}$ uniquely and $\boldsymbol{h}$ up to an additive constant $[c_1\mathbf{e}, c_2\mathbf{e}]$ with $c_1, c_2 \in R$ $(c_1 < c_2)$.

3 Asymptotic properties of $v_T(f)$

In this section we study the asymptotic behavior of $v_T(f)$ as $T \to \infty$ under Assumption A.

Throughout this section, we assume that Assumtion A holds.

For any $g \in C(R)^n$ and $f \in F$, the sequence $\{v_T(f, g), T \geq 0\}$ is defined as follows:

$$v_0(f, g) := g \qquad \text{and}$$

$$v_T(f, g) := \{r(f) + Q_1 r(f) + \cdots + Q_1 \cdots Q_{T-1} r(f) + Q_1 \cdots Q_T g \mid Q_i \in \mathcal{Q}(f), i = 1, \cdots, T\} \quad (t \geq 1). \tag{3.1}$$

Lemma 3.1 For any $g \in C(R_+)^n$ and $f \in F$, the sequence $\{v_T(f, g)\}$ satisfies that

$$v_T(f, g) = L(f) v_{T-1}(f, g) \quad (T \geq 1). \tag{3.2}$$

Since the solutions $\underline{v}$, $\overline{v}$, h of (2.6)-(2.8) in Section 2 are depending on $f \in F$, we will denote them respectively by $v(f) = [\underline{v}(f)\mathbf{e}, \overline{v}(f)\mathbf{e}]$ and $h(f) = [\underline{h}(f), \overline{h}(f)]$.

Lemma 3.2 For any $f \in F$, it holds that

$$v_T(f, h(f)) = T v(f) + h(f) \quad \text{for all } T \geq 0. \tag{3.3}$$

The following theorem is concerned with the asymptotic properties of $v_T(f)$ as $T \to \infty$.

Theorem 3.1 For any $f \in F$, there exists $c_1,\ c_2,\ c_1',\ c_2' \in R\ \ (c_1' \leq c_1,\ c_2' \leq c_2)$ such that

$$[(T\underline{v}(f) + c_1)\mathbf{e}, (T\overline{v}(f) + c_2')\mathbf{e}] \subset v_T(f) \subset [(T\underline{v}(f) + c_1')\mathbf{e}, (T\overline{v}(f) + c_2)\mathbf{e}] \quad \text{for all } T \geq 0, \tag{3.4}$$

where $[a, b] = \emptyset$, if $a > b$.

Corollary 3.1 For any $f \in F$, it holds

(i) $$v(f) = [\underline{v}(f)\mathbf{e}, \overline{v}(f)\mathbf{e}]$$

and

(ii) $$\underline{v}(f)\mathbf{e} = \underline{Q}^* r(f), \quad \overline{v}(f)\mathbf{e} = \overline{Q}^* r(f)$$

for any $\underline{Q} \in \underline{\mathcal{Q}}(f, \underline{h})$ and $\overline{Q} \in \overline{\mathcal{Q}}(f, \overline{h})$, where $Q^* = \lim\limits_{T\to\infty} \dfrac{1}{T} \sum\limits_{t=0}^{T-1} Q^t$. for $Q \in \mathcal{Q}(f)$.

4 Average-opimal policies

In this section, we give the existence theorem of an average-optimal policy under Assumption A.

Let $\mathbf{q}(i,a) := \langle \underline{q}(\cdot|i,a), \overline{q}(\cdot|i,a)\rangle$ for each $i \in S$ and $a \in A$.

For each $i \in S$ and $f \in F$, denote by $\underline{G}(i,f)$ the set of $a \in A$ for which

$$\underline{v}(f) + \underline{h}(f)_i < r(i,a) + \min_{q \in \mathbf{q}(i,a)} \sum_{j=1}^{n} q(j|i,a)\underline{h}(f)_j,$$

where $\underline{v}(f)$ and $\underline{h}(f) = (\underline{h}(f)_1, \cdots, \underline{h}(f)_n)$ is solution of (2.6).

Let $g \in F$ be such that $g(i) \in \underline{G}(i,f)$ for any i with $\underline{G}(i,f) \neq \emptyset$ and $g(i) = f(i)$ for any i with $\underline{G}(i,f) = \emptyset$.

Then, we have the following.

Lemma 4.1 For any f with $\underline{G}(i,f) \neq \emptyset$ for some $i \in S$, $\underline{v}(f) < \underline{v}(g)$.

The following lemma is proved from the idea of policy improvement (cf.[8]).

Lemma 4.2 The left side optimality equations (4.1) below determine $\underline{v}^*$ uniquely and $\underline{h} \in R^n$ up to an additive constant.

$$\underline{v}^* + \underline{h}_i = \max_{a \in A} \left(r(i,a) + \min_{q \in \mathbf{q}(i,a)} \sum_{j=1}^{n} q(j|i,a)\underline{h}_j \right) \quad (1 \leq i \leq n). \tag{4.1}$$

Let, for each i $(1 \leq i \leq n)$,

$$A_i := arg \max_{a \in A} \left(r(i,a) + \min_{q \in \mathbf{q}(i,a)} \sum_{j=1}^{n} q(j|i,a)\underline{h}_j \right).$$

For each $i \in S$ and $f \in F$ with $f(i) \in A_i$ for all $i \in S$, denote by $\overline{G}(i,f)$ the set of $a \in A_i$ for which

$$\overline{v}(f) + \overline{h}(f)_i < r(i,a) + \max_{q \in \mathbf{q}(i,a)} \sum_{j=1}^{n} q(j|i,a)\overline{h}(f)_j,$$

where $\overline{v}(f)$ and $\overline{h}(f) = (\overline{h}(f)_1, \cdots, \overline{h}(f)_n)$ is a solution of (2.7).

Using $\overline{G}(i,f)$ instead of $\underline{G}(i,f)$ and applying the same way as the proof of Lemma 4.2, we can prove the following.

Lemma 4.3 The right side optimality equations (4.2) below determine $\overline{v}^*$ uniquely and $\overline{h} \in R^n$ up to an additive constant.

$$\overline{v}^* + \overline{h}_i = \max_{a \in A_i} \left(r(i,a) + \max_{q \in \mathbf{q}(i,a)} \sum_{j=1}^{n} q(j|i,a)\overline{h}_j \right) \quad (1 \leq i \leq n). \tag{4.2}$$

Let, for each i $(1 \leq i \leq n)$,

$$A_i^* := arg \max_{a \in A_i} \left(r(i,a) + \max_{q \in \mathbf{q}(i,a)} \sum_{j=1}^{n} q(j|i,a)\overline{h}_j \right).$$

Then we have the following Theorem.

Theorem 4.1 Let f^* be any policy with $f^*(i) \in A_i^*$ for all $i \in S$. Then f^* is average-optimal and $v(f^*) = [\underline{v}^* \mathbf{e}, \overline{v}^* \mathbf{e}]$.

Here we shall give a numerical example which illustrates Theorem 4.1.

For simplicitity, let $\underline{q}_{ij}^a := \underline{q}(j|i,a)$, $\overline{q}_{ij}^a := \overline{q}(j|i,a)$ and $r(a) := (r(1,a), r(2,a))$. Consider the following controlled Markov set-chain model :

$S = \{1,2\}, A = \{1,2\}$,

$$(\underline{q}_{ij}^1) = \begin{pmatrix} \frac{1}{3} & \frac{1}{3} \\ \frac{1}{3} & \frac{1}{3} \end{pmatrix}, (\overline{q}_{ij}^1) = \begin{pmatrix} \frac{2}{3} & \frac{1}{2} \\ \frac{1}{2} & \frac{2}{3} \end{pmatrix},$$

$$(\underline{q}_{ij}^2) = \begin{pmatrix} \frac{2}{5} & \frac{2}{5} \\ \frac{2}{5} & \frac{2}{5} \end{pmatrix}, (\overline{q}_{ij}^2) = \begin{pmatrix} \frac{1}{2} & \frac{3}{5} \\ \frac{3}{5} & \frac{1}{2} \end{pmatrix},$$

$r(1) = (1,2)$ and $r(2) = (1,2.1)$.

Then, the equation (4.1) which $\underline{v}^*$ and $\underline{h} = (\underline{h}_1, \underline{h}_2)'$ is given as follows :

$$\underline{v}^* + \underline{h}_1 = \max \begin{cases} 1 + \min\{\frac{2}{3}\underline{h}_1 + \frac{1}{3}\underline{h}_2, \frac{1}{2}\underline{h}_1 + \frac{1}{2}\underline{h}_2\} \\ 1 + \min\{\frac{2}{5}\underline{h}_1 + \frac{3}{5}\underline{h}_2, \frac{1}{2}\underline{h}_1 + \frac{1}{2}\underline{h}_2\}, \end{cases}$$

$$\underline{v}^* + \underline{h}_2 = \max \begin{cases} 2 + \min\{\frac{1}{3}\underline{h}_1 + \frac{2}{3}\underline{h}_2, \frac{1}{2}\underline{h}_1 + \frac{1}{2}\underline{h}_2\} \\ 2.1 + \min\{\frac{3}{5}\underline{h}_1 + \frac{2}{5}\underline{h}_2, \frac{1}{2}\underline{h}_1 + \frac{1}{2}\underline{h}_2\}, \end{cases}$$

After a simple calculation, the solution of the above with $\underline{h}_1 = 0$ becomes that $\underline{v}^* = 1.5$ and $\underline{h} = (0,1)'$. Also, we easily find $A_1 = \{2\}$ and $A_2 = \{1,2\}$.

Similarly, by solving the equation which $\overline{h}_1 = 0$, we get $\overline{v}^* = 23/14, \overline{h} = (0, 15/14)', A_1^* = \{2\}$ and $A_2^* = \{1\}$. So, by Theorem 4.1, f^* with $f^*(1) = 2$ and $f^*(2) = 1$ is average-optimal and $v(f^*) = [(3/2)\mathbf{e}, (23/14)\mathbf{e}]$

References

[1] Bather, J. (1973) Optimal decision procedures for finite Markov chains. Part II : Commumicating systems. *Adv. Appl. Prob.* 5, pp.521-540.

[2] Blackwell, D. (1962) Discrete dynamic programing. *Ann. Math. Statist.* 33, pp.719-726.

[3] Hartfiel, D. J. (1981) On the limiting set of stochastic product $xA_1 \cdots A_m$. *Proc. Amer. Math. Soc.* 81, pp.201-206.

[4] Hartfiel, D. J. (1991) Component bounds for Markov set-chain limiting sets. *J. Statist. Comput. Simul.* 38, pp.15-24.

[5] Hartfiel, D. J. (1993) Cyclic Markov set-chains. *J. Statist. Comput. Simul.* 46, pp.145-167.

[6] Hartfiel, D. J. and Seneta, E. (1994) On the theory of Markov Set-chains. *Adv. Appl. Prob.* 26, pp.947-964.

[7] Hinderer, K. (1970) *Foundations of Non-Stationary Dynanic Programming with Discrete Time Parameter.* Springer-Verlag, New York.

[8] Howard, R. (1960) *Dynamic Programming and Markov processes.* MIT Press, Cambridge MA.

[9] Kemeny, J. G. and Snell, J. L. (1960) *Finite Markov-Chains.* Van Nostrand, New York.

[10] Kurano, M. , Song, J. , Hosaka, M. and Huang, Y. Controlled Markov Set-Chains with Discounting. *To appear in J.Appl.Prob.* (1998).

[11] Nenmaier, A. (1984) New techniques for the analyses of linear interval equations. *Linear Algebra Appli.* 58, pp.273-325.

[12] Puterman, M. L. (1994) *Markov decision processes: Discrete Stochastic Dynamic Programming.* Jhon Wiley & Sons, INC.

[13] Seneta, E. (1981) *Nonnegative Matrices and Markov Chains.* Springer-Verlag, New York.

[14] Song, J. , Hosaka, M. and Kurano, M. A Span Seminorm Approach to Controlled Markov Set-chains. (Preprint).